ELSEVIER

Displays 20 (1999) 93–110

DISPLAYS

Image quality metric based on multidimensional contrast perception models

A.M. Pons*, J. Malo, J.M. Artigas, P. Capilla

Departament d'Óptica, Facultat de Física, Universitat de València, C/Dr. Moliner 50, Burjassot, 46100 València, Spain

Received 20 November 1998; received in revised form 21 April 1999; accepted 26 April 1999

Abstract

The procedure to compute the subjective difference between two input images should be equivalent to a straightforward difference between their perceived versions, hence reliable subjective difference metrics must be founded on a proper perception model.

For image distortion evaluation purposes, perception can be considered as a set of signal transforms that successively map the original image in the spatial domain into a feature and a response space. The properties of the spatial pattern analyzers involved in these transforms determine the geometry of these different signal representation domains.

In this work the general relations between the sensitivity of the human visual system and the perceptual geometry of the different representation spaces are presented. This general formalism is particularized through a novel physiological model of response summation of cortical cells that reproduce the psychophysical data of contrast incremental thresholds. In this way, a procedure to compute subjective distances between images in any representation domain is obtained.

The reliability of the proposed scheme is tested in two different contexts. On the one hand, it reproduces the results of suprathreshold contrast matching experiences and subjective contrast scales (Georgeson and Shackleton, Vision Res. 34 (1994) 1061–1075; Swanson et al., Vision Res. 24 (1985) 63–75; Cannon, Vision Res. 19 (1979) 1045–1052; Biondini and Mattiello, Vision Res. 25 (1985) 1–9), and on the other hand, it provides a theoretical background that generalizes our previous perceptual difference model (Malo et al., Im. Vis. Comp. 15 (1997) 535–548) whose outputs are linearly related to experimental subjective assessment of distortion.

Keywords: Subjective image distortion metric; Contrast response summation model; Information allocation by the human viewer

1. Introduction

The design of image difference measures, reproducing the subjective distance judgements by an average observer, has both applied and academic interest.

On the one hand, such perceptually correlated difference measures have a key importance in the design of signal processing algorithms in visual communication systems to be evaluated by human observer [1]. Observer-oriented imaging systems, as multimedia broadcasting, medical image database or surveillance devices for target detection, should minimize the *perceptual distortion* of the output. The measure of distortion is of central importance in image and video coding techniques, motion estimation procedures and adaptive noise removal for image enhancement. Hence, they can be perceptually optimized if proper subjective metric is used to guide their design.

On the other hand, the reproduction of the observer's opinion in a complex task such a distortion assessment is a challenging problem that can be used to discriminate between alternative perception models. For instance, in the related context of color difference assessment, the empirical ad hoc color difference formulae, such as *CIELab*, are founded on color appearance models to obtain appropriate results [2]. Moreover, any proposed color vision model must give rise to a reliable distance expression to be valid.

Measuring *subjective difference between input images* defined in the spatial domain should correspond to the evaluation of the *difference between the perceived images*. This implies that general purpose subjective metrics should emerge from models including the behavior of the mechanisms underlying visual perception.

To this end, signal processing in the visual pathway can be considered as a set of signal mappings from the original spatial domain to a visual mechanism's response domain. The whole transform can be viewed as a two-stage process:

* Corresponding author.
E-mail address: jesus.malo@uv.es (J. Malo)

0141-9382/99/$ - see front matter
PII: S0141-9382(99)00009-8

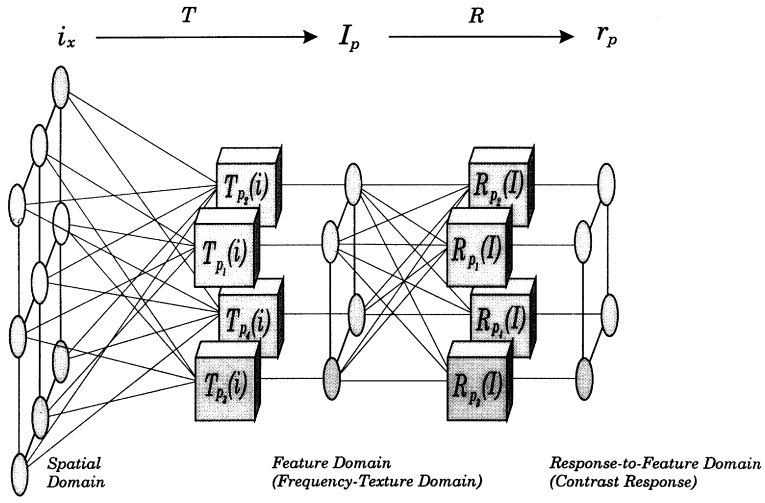

Fig. 1. Scheme of the two-stage process of low-level perception. The original signal in the spatial domain, i_x, is mapped into a transformed domain of features p. In the second stage, R, a number of specialised mechanisms non-linearly responds to the amplitude I_{p_i} of the different features p_i in the input image, giving rise to the response representation, r_p.

in the first place, a low-level feature extraction process is carried out and then, a response-to-features transform is performed [3–7] (see Fig. 1).

This low-level two-stage process considered here is the first bottle neck to the visual information flow in the image perception. Hence, it can be used in the general purpose image-independent applications to define the upper bound of image distortion to be noticeable. Obviously, higher-level interactions between the responses occur, but they are outside the scope of this work due to their complexity and their scene content dependent nature.

It is commonly assumed that the early feature extraction process (including the local extraction of texture and edge orientation at different scales) is performed through the application of a set of Gabor or wavelet band-pass filters [4–11]. This multichannel filtering process maps the original signal to a local frequency-texture domain [12–13].

The second stage is performed through a set of non-linear mechanisms tuned to certain regions of the feature space (tuned to certain frequency bands and spatial positions). These mechanisms respond to the *strength* (amplitude or contrast) of a given feature in the input signal [3–6].

Taking into account that a given local frequency transform has been computed at the first perception stage, the psychophysical and physiological experimentation try to elucidate the non-linear response of the mechanisms tuned to the different basis functions of such a paradigm [14–16]. The individual cells of the Magno and Parvocellular pathways [17–18] are the actual biological hardware of these abstract analyzers. The models of summation of the individual responses of these cells have to reproduce the global psychophysical properties.

The different sensitivity, relative gain, and non-linearity of the spatial analyzers emphasize some regions of the feature space while reducing the perceptual importance of other regions. Hence, the *perceptual geometry* of any representation domain (the procedure to compute subjective differences between two images in that domain) should be computable from the geometry in other domain and the properties of the transform mechanisms.

Therefore, if reliable data or a successful model of the mechanism's response is attached to the proper geometric formalism, a complete scheme to compute image differences in the desired representation domain will be obtained.

Within this framework, two main contributions are presented in this paper:

(a) Some general relations between the perceptual geometry of the different representation domains are proposed. An explicit expression of the perceptual metric tensor as a function of the gradients of the contrast response and the feature extraction transform is derived. It is shown how this description of the perceptual geometry can be used with a wide class of contrast perception models or data.
(b) As a practical application, the general formalism is reduced to a particular image distortion measure through a novel physiological model of contrast response summation of cortical cells. The proposed response model takes into account the joint contributions of the parvo and magnocellular pathways of the human visual system with a frequency-dependent pooling exponent and reproduces the psychophysical contrast incremental threshold data of sinusoidal gratings.

The reliability of the whole scheme (metric formalism, together with a particular physiological response model) is tested in two different contexts. On the one hand, it reproduces the results of suprathreshold contrast matching experiences and subjective contrast scales [19–22], and on the other hand, it provides a unified theoretical background to some linear and non-linear transform-based perceptual difference models [23–27]. It also generalizes our

previously reported distortion assessment algorithm [28] whose outputs are linearly related to experimental subjective assessment of impairment.

The structure of this paper is as follows: in Section 2, the general formalism relating the sensitivity and the perceptual geometry of a given representation domain with the transforms and non-linearities of the visual mechanisms is presented. The physiological model is developed and its psychophysical consequences outlined in Section 3. In Section 4, using the model to particularize the formalism, the non-uniform sensitivity of the HVS in a particular frequency-and-contrast domain and the corresponding expression to compute subjective image differences are presented. Section 5 shows how the complete scheme reproduces experimental data of contrast perception and subjective quality assessment. Some final remarks are made in Section 6.

2. Sensitivity of the visual system and perceptual geometry in different representation spaces

In this section, some usual psychophysical terminology is geometrically formalized, and the relations between the methods to compute perceptual image differences and the experimental data or response models are derived. These relations will be used to obtain the perceptual metrics from the non-linear contrast response model of Section 3. The same definitions could be applied to direct psychophysical data of incremental threshold in any transform domain. The formalism can deal with non-linear feature extraction transforms and with complex post-transform masking schemes.

2.1. Basic terminology

2.1.1. System response and representation domains

As stated above (Fig. 1), the standard formulation of contrast perception models [3–7] assume that the system response to a given input image is the result of changing the representation domain through a transform, T, followed by a non-linear response transform, R:

$$\text{Response} \equiv i_x \in R^m \xrightarrow{T} I_p \in R^n \xrightarrow{R} r_p \in R^n. \tag{1}$$

The *spatial domain representation* of the signal is the input luminance function at m points in the image plane, $i_x = (i_{x_1}, i_{x_2}, i_{x_3}, \ldots i_{x_m})$. The *feature space representation* of the signal is the transformed function $I_p = (I_{p_1}, I_{p_2}, I_{p_3}, \ldots I_{p_n})$, where each parameter p_i represent a particular basis function of the transform T. I_{p_i} gives the *strength* of the feature p_i in the input signal. The *response or neural representation* of a signal is the point $r_p = (r_{p_1}, r_{p_2}, r_{p_3}, \ldots, r_{p_n})$ obtained from I_p after the application of the response functions $R_{p_i}(I)$ of the detectors tuned to each feature p_i. Note that, in general, this formulation may include changes in the dimensionality of the representation, and masking between features. Each response component, r_{p_i}, is a function of every component of the signal in the feature space representation, i.e. the response of the detector of the feature p_i not only depends on the *contrast* of the basis function p_i in the input signal, I_{p_i}, but it may also have contributions from the other components: $r_{p_i} = R_{p_i}((I_{p_1}, I_{p_2}, I_{p_3}, \ldots, I_{p_n}))$.

2.1.2. Perceptual distance between images

Given two input images represented in the response space by r_p and $r_p + \Delta r_p$, respectively, the perceptual difference between them must be a function of the unidimensional differences, Δr_{p_i}, i.e. a function of the response of the individual detectors tuned to each feature of the transform domain. A *β-order* norm of the difference vector is widely used to take into account the contribution of each individual detector [29–30]:

$$D(r, r + \Delta r) = \|\Delta r\|_\beta = \left(\sum_{p_i, p_j} (\Delta r^{p_j})^{\beta/2} W(r)^{p_i}_{p_j} (\Delta r_{p_i})^{\beta/2} \right)^{1/\beta}, \tag{2}$$

where β is the so-called summation index. This expression allows mutual interactions between the different specialized detectors through the crossed terms $p_i \neq p_j$.

A perceptual hierarchy between the different dimensions (features, textures or frequencies) at a given point r according to the weights of the *metric matrix*, $W(r)$, which control the contribution of each unidimensional distortion to the global perception of distortion can be established. The perceptually relevant features of the signal are those with a significant weight. Intuitively, it can be seen that there is a close relationship between metric and sensitivity.

As we shall see, point-dependent metrics can be used to describe domains where the sensitivity is non-uniform.

While the metric controls the relative importance of each variation in the detector response to the global perception of change in the input, the summation index controls *how to sum* these different contributions. Basically, three different behaviors are obtained as a function of the summation index [30]: linear behavior ($\beta = 1$), vectorial behavior ($\beta = 2$), and peak detection behavior ($\beta \to \infty$).

Considering a constant value of distance in Eq. (2), one has the equation of equal distortion locus at a point r. In this superquadric equation [31], the values of the metric determine the relative length and orientations of the principal axis of the quadric whereas the β parameter determines the convexity of the loci.

The different perceptual models agree in the uniformity of the response space, and differ in the pooling exponent, β, and in the expression of the non-linear transform R. The simpler models [3,29] assume that R functions are computable from the experimental data about amplitude JNDs of isolated basis functions. In the more recent and complex models [6], the non-linearity of the response to each basis function depend, to a certain extent, on the amplitude of the

other frequency coefficients. The value of the exponent β is fixed to 2 in some schemes [3] (or to 2.4 in Ref. [9]), assuming the so-called ideal observer approximation [6,29], while a greater value ($\beta \sim 4$) is used in other special cases [6].

For mathematical convenience, a vectorial summation, $\beta = 2$, will be assumed here in every representation space. This constraint does not substantially affect the consequences obtained from the distortion measure because, as we shall see, our interest will be focused on the relative length of the principal axis of the locus at a given point and on the relative volume of the loci at different points of the domain, which exclusively depend on the metric, and not on the summation index.

It will be assumed that in the neural representation space, all unidimensional increments are equally weighted and there is no cross-correlation between the detector response in the distortion computation, i.e. euclidean metric will be assumed in the response domain [3]. This assumption is consistent with the equivalence of perception and neural representation at this level: once in the response representation, all regions and directions of the domain should have the same perceptual importance. Any way, the assumption of an identity metric for the response space does not imply any constraint to the final model because any non-uniformity or interaction between the responses could be included as interdependencies between the components of I_p and non-linearities in the functions $R_p(I)$.

In this way, the perceptual difference between two patterns in the considered domains (response domain, feature domain and spatial domain) can be expressed as follows:

$$D^2(r, r + \Delta r) = \Delta r^{\mathrm{T}} \Delta r, \tag{3}$$

$$D^2(I, I + \Delta I) = \Delta I^{\mathrm{T}} W \Delta I, \tag{4}$$

$$d^2(i, i + \Delta i) = \Delta i^{\mathrm{T}} w \Delta i. \tag{5}$$

The invariance of the perceptual difference with regard to the representation domain is the key to obtain the relationship between the properties of the mechanisms of the visual pathway and the metric in each domain. In Section 2.3, it is shown how the metrics in the spatial domain and in the feature space can be obtained from the non-linear response of the feature detectors.

2.1.3. Incremental thresholds and discrimination boundaries

The incremental thresholds in the neighborhood of a given point, i_0, I_0, or r_0, are the variations $\Delta i(i_0)$, $\Delta I(I_0)$, or $\Delta r(r_0)$, to be added to the given patterns to make the result just discriminable from the original signal.

The discrimination boundaries in the neighborhood of a given point are the loci of points that are at just noticeable distance (JND) from the central point.

Assuming the distance expressions (3)–(5), the discrimination boundaries in the response space will be spheres of uniform radius all over the domain, while in the other representation spaces, their shape and volume will depend on the transforms T and R.

2.1.4. Sensitivity and resolution in a representation domain

The sensitivity of the system at a given point of a representation domain is proportional to the inverse of the euclidean volume of the ellipsoidal discrimination boundary at that point. The sensitivity is high in the regions where a small variation in the signal is enough to put the representative point beyond the discrimination boundary. This definition is equivalent to a definition of resolution. The system is named to have more sensitivity in the regions where the resolution (number of discriminable patterns per unit of area) is bigger.

As the volume of an ellipsoid defined by a certain quadratic form is proportional to the product of the eigen-values of the inverse of this matrix, the sensitivity at the different representation spaces can be defined by:

$$S(r) = \sqrt{\det(W(r))} = 1, \tag{6}$$

$$S(I) = \sqrt{\det(W(I))}, \tag{7}$$

$$s(i) = \sqrt{\det(w(i))}. \tag{8}$$

A definition of the sensitivity exclusively based on the volume of the discrimination loci does not emphasize the fact that the discrimination may be dimension-dependent. It may depend on the feature (e.g. it may depend on the frequency).

This is particularly important if linear transformations and responses are considered, such as for example in the classical models that use Fourier transforms and linear filtering [32], or more recently proposed Gabor or Wavelet transforms and linear filtering [11]. In these cases, the ellipsoidal discrimination boundaries are scaled and rotated in the same way all over the domain, thereby keeping a constant volume for every point of the representation. In such cases, it is desirable to have a definition of sensitivity *in each dimension* p_i, $S_{p_i}(r)$, $S_{p_i}(I)$ or $S_{x_i}(i)$, inversely proportional to the incremental thresholds in the direction of the axis p_i of the representation.

$$S_{p_i}(r) = \sqrt{W(r)_{p_i}^{p_i}} = 1, \tag{9}$$

$$S_{p_i}(I) = \sqrt{W(I)_{p_i}^{p_i}}, \tag{10}$$

$$S_{x_i}(i) = \sqrt{w(i)_{x_i}^{x_i}}. \tag{11}$$

In linear characterizations of the response in feature domains with frequency meaning, the sensitivity, $S_{p_i}(I)$, is the classical *Contrast Sensitivity Function* [32], valid for near-threshold amplitudes ($I_{p_i} \ll$).

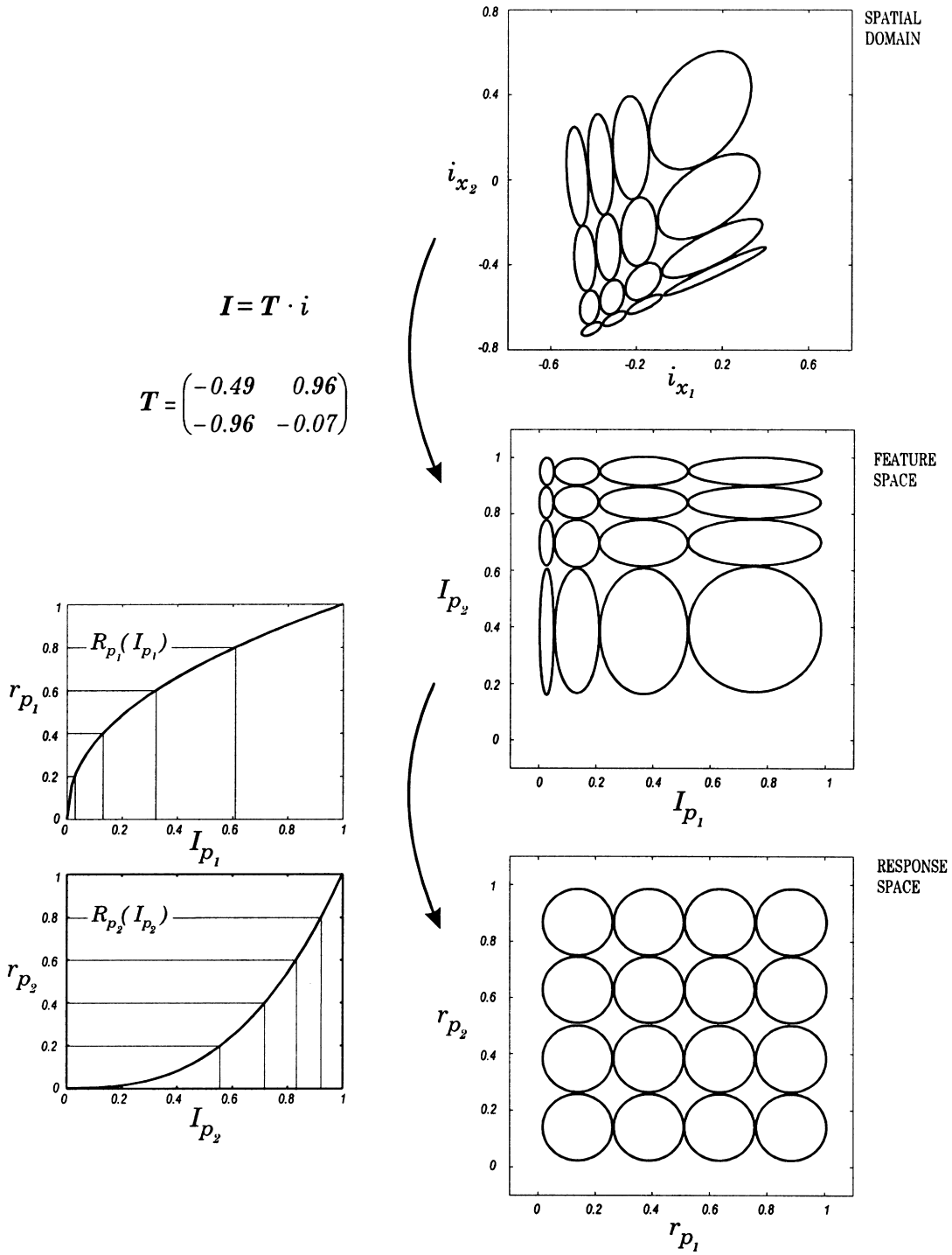

Fig. 2. Two-dimensional example of the effect of the transforms T and R in the metric and in the shape of the discrimination boundaries according to expressions (16) and (18). The discrimination boundaries at a set of corresponding points in the different representation spaces are shown on the right. The expressions and the shape of the transforms used in this example are shown on the left.

2.2. Response, perceptual metric and sensitivity from incremental thresholds

Assuming that certain local frequency transform has been carried out, the psychophysical experimentation use the basis functions of the transform to measure the incremental thresholds in the axis directions of the feature space, varying the amplitude of the input stimulus until the observer detects the changes.

In this section, we show how to use these experimental data to compute the system characterization: the sensitivity

functions and the perceptual metric of the representation spaces.

2.2.1. System response and feature incremental thresholds

Consider a linear approximation of the response function, $R(I)$, in a neighborhood of a point, I_0:

$$R(I_0 + \Delta I) = R(I_0) + \nabla R(I_0)\Delta I. \tag{12}$$

In this case, the variation of the response when varying the signal in the feature space is:

$$\Delta r = \nabla R(I_0)\Delta I, \tag{13}$$

where

$$\nabla R(I_0) = \left.\begin{pmatrix} \frac{\partial R_{p_1}(I)}{\partial I_{p_1}} & \cdots & \frac{\partial R_{p_1}(I)}{\partial I_{p_n}} \\ \vdots & & \vdots \\ \frac{\partial R_{p_n}(I)}{\partial I_{p_1}} & \cdots & \frac{\partial R_{p_n}(I)}{\partial I_{p_n}} \end{pmatrix}\right|_{I = I_0}.$$

Considering the JND and substituting the increment, $\Delta r = \nabla R(I_0)\Delta I$, in the expression of the difference in the response domain (Eq. (3)), we have an equation that relates the experimental incremental thresholds in the feature space with the gradient of the response function:

$$\Delta I^{\mathrm{T}} \nabla R^{\mathrm{T}} \nabla R \Delta I - \mathrm{JND} = 0. \tag{14}$$

With the proper experimental measures of incremental thresholds, this system of equations can be solved numerically as an optimization problem, searching for the parameters ∇R that fulfill the constraint (14) [33].

As proposed in a color vision context [34], this differential characterization of the detection mechanisms, through the local gradient $\nabla R(I)$, is really powerful if the response representation is additive, not necessarily linear. If the response is additive, assuming a certain value of the response of the system at a given point, such as the origin, the response of the system at any point can be obtained by just adding the differential variations obtained from Eq. (14).

$$r(I) = r(I_0) + \int_{I_0}^{I} \nabla R(I')\, \mathrm{d}I'. \tag{15}$$

In the linear case, this expression could be further simplified as the gradient of the response would be constant.

2.2.2. Perceptual metric and non-linearities in the response

Comparing the expressions (4) and (14) in the case of incremental thresholds and JNDs, it is clear that the gradient of the response function has the role of metric in (14), hence

$$W = \nabla R^{\mathrm{T}} \nabla R. \tag{16}$$

This is consistent with the tensorial character of the metric. The general definition of a tensor establishes that given a generic co-ordinate change, h, the coefficients of a tensor G defined in the transformed domain are related to their expression, g, in the original domain through [35]:

$$g = \nabla h^{\mathrm{T}} G \nabla h \tag{17}$$

which reduces to expression (16), taking the identity as metric in the neural representation as assumed above.

Reasoning in the same way, in the case of the transform from the spatial domain to the feature representation, it holds that

$$w = \nabla T^{\mathrm{T}} \nabla R^{\mathrm{T}} \nabla R \nabla T \tag{18}$$

From expression (16), it can be stated that perceptual distances in the feature domain fully rely on the non-linearities of the response R. The perceptual effect of the amplitude variations in the feature space, and the volume of the discrimination ellipsoids will depend on the slope of the response function. It will be small in smooth response areas while it will be bigger in the rapidly changing response areas. If the response is linear, the gradient is constant and this implies a point invariant metric (and uniform sensitivity).

The orientation of these discrimination boundaries will also depend on the characteristics of the response function. If the response of the detector of each feature p_i exclusively depends on the amplitude of that feature in the input signal, I_{p_i}, then the gradient matrices, ∇R, and the metric, W, will be diagonal, and the discrimination ellipsoids will be aligned with the axis of the representation. If some interaction exists between features, Eq. (16) implies that the metric will be non-diagonal and the ellipsoids will be rotated accordingly.

Fig. 2 shows an explicit example of a system with 2D inputs (only two pixels in the spatial domain) and a 2D feature space. In this case, the feature extraction transform is just a linear change of coordinates (such a Fourier transform or a projection over a basis of wavelet functions). The detectors tuned to each feature have different non-linear response without masking between the features (such as in independent multichannel schemes).

The discrimination boundaries at a set of corresponding points in the different representation spaces are shown in the figures.

As stated above, an euclidean metric is assumed in the response domain. It can be seen how the non-linearities of the response functions, R, determine the non-uniformities of the feature domain as expressed in Eq. (16). As there is no masking between the features, ∇R is diagonal, and then, the discrimination regions are aligned with the axis.

The non-linear responses imply that the gradient is point-dependent, and the shape and the volume of the discrimination regions become non-uniform in the feature domain. Note how the regions with high slope response (low values of I_{p_1} and high values of I_{p_2}) are *dilated* in the response representation emphasizing their perceptual importance. However, the regions with a smooth response (high values of I_{p_1} and low values of I_{p_2}) are *compressed* by the R

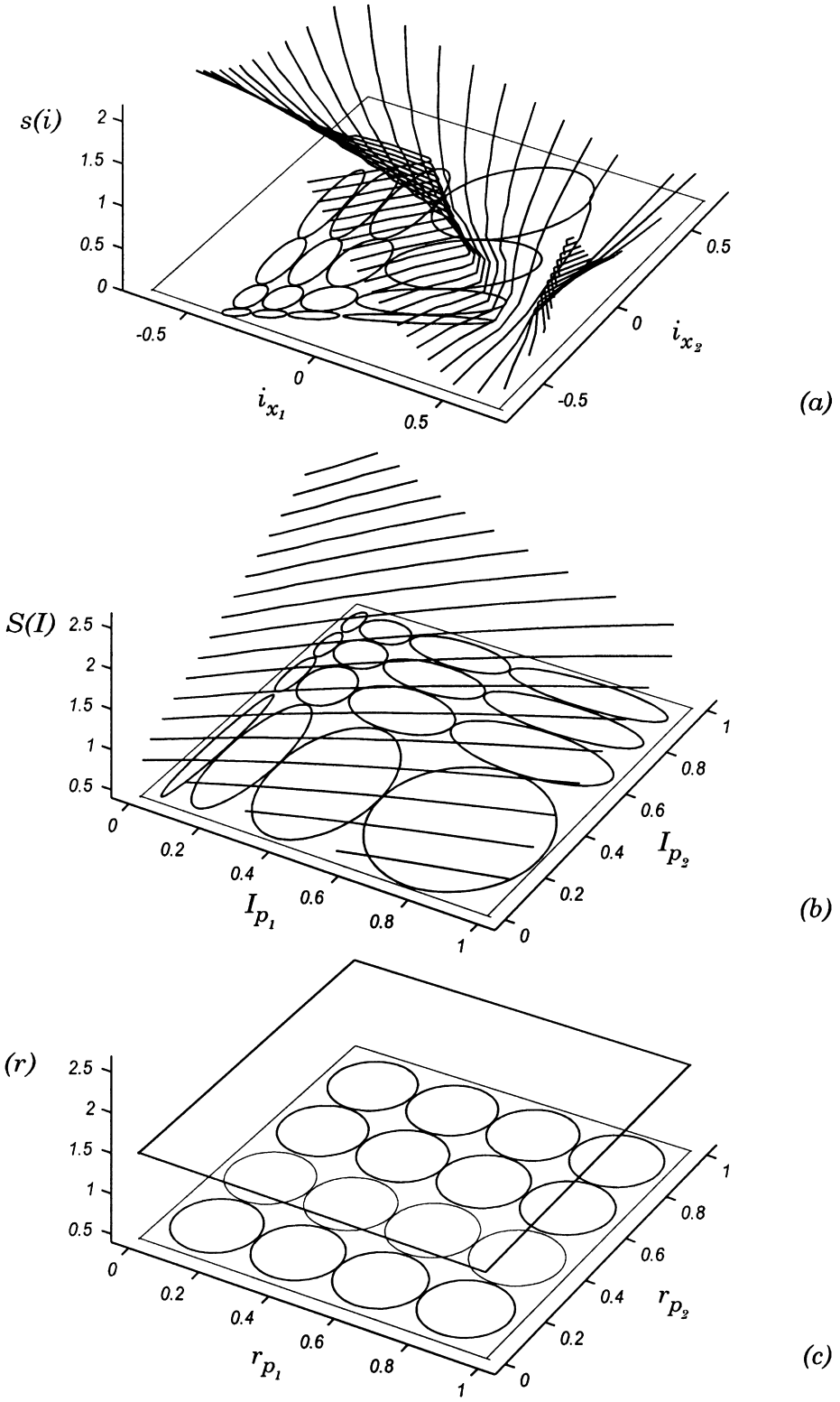

Fig. 3. Sensitivity of the system described in the example of Fig. 2 in the different representation domains.

transform, giving relatively low volume regions in the final representation.

As the T transform is linear, the discrimination regions in the space domain are uniformly scaled and rotated to give the ellipsoid regions of the feature domain.

2.2.3. Sensitivity and non-linearities in the response

A straightforward substitution of the expressions of the perceptual metric in the definitions (6)–(11) gives the relation between the sensitivity at any point of the representation domains as a function of the gradients of the response transforms:

$$S(r) = \sqrt{\det(I)} = 1, \tag{19}$$

$$S(I) = \sqrt{\det(\nabla R(I)^{\mathrm{T}} \nabla R(I))}, \tag{20}$$

$$s(i) = \sqrt{\det(\nabla T(i)^{\mathrm{T}} \nabla R(T(i))^{\mathrm{T}} \nabla R(T(i)) \nabla T(i))} \tag{21}$$

in the same way, the sensitivity in each dimension is:

$$S_{p_i}(r) = \sqrt{I_{p_i}^{p_i}} = 1, \tag{22}$$

$$S_{p_i}(I) = \sqrt{\nabla R_{p_i}^{p_k} \nabla R_{p_k}^{p_i}}, \tag{23}$$

$$s_{x_i}(i) = \sqrt{\nabla T_{x_i}^{p_k} \nabla R_{p_k}^{p_i} \nabla R_{p_l}^{p_m} \nabla T_{p_m}^{x_i}}, \tag{24}$$

where summation is assumed over repeated indices.

Fig. 3 shows the sensitivity surfaces in the different representation domains for the example of Fig. 2. The assumption of euclidean metric in the response domain implies uniform sensitivity in this domain. The non-linearities in the response functions give rise to dilations and compressions of the regions of the feature space when transformed by R. This implies high sensitivity in the dilated regions and low sensitivity in the compressed ones. The linear feature extraction transform T rotate and scale the sensitivity surface, maintaining the relative differences in sensitivity between corresponding points in the spatial domain and in the feature domain. The surface shown in Fig. 3(b) corresponds to the surface in Fig. 3(a) over the considered discrimination ellipsoids: the other part of the surface is outside of the considered range in Fig. 3(b) and (c).

2.3. Derivation of simple perceptual metrics from linear perception models

The formalism presented in the previous sections reproduce the metrics described in the papers of Nill [23], Granger [24] or in the review of Barten [25], when the low-level processing of the visual system is modeled as a linear, space invariant system. The metrics that include the CSF or other qualitatively similar Fourier filter show high correlation in the opinion of the observer [28]. Assuming a linear behavior, the system can be described by a weighting function over the coefficients of a linear transform such as the Fourier transform [32] or a wavelet transform [11]. This coefficient-selective weighting function has been interpreted as the sensitivity of the system in each dimension of the transformed domain, S_{p_i} (the CSF in the Fourier domain or its equivalent in other basis domain).

If no interaction between features is considered, which is a good approximation when using wide band channels such as in a Gabor or wavelet transform [4,8–10], the transform R will be a diagonal matrix containing the linear weights (the sensitivity function) in the diagonal: $R_{p_i}^{p_i} = S_{p_i}$.

Under these assumptions, the transforms T and R are linear, and their gradients equal to the transform matrices. Hence, the expressions for the metrics in the frequency and the spatial domain are:

$$W = R^{\mathrm{T}} R, \tag{25}$$

$$w = T^{\mathrm{T}} R^{\mathrm{T}} R T, \tag{26}$$

where the orthogonal matrix T contains the basis functions [36], and the non-zero coefficients of the diagonal matrix R are given by the sensitivity weights.

Note that the perceptual relevance of each feature in the frequency domain, the diagonal of the metric W, is given by the square of the sensitivity function. Moreover, the metric in the spatial domain (the perceptual effect of the surrounding on a given point) is given by the autocorrelation of the impulse response (the inverse transform of the square of the sensitivity filter function) which is point invariant.

In this case, the perceptual difference between two images, i and i', is:

$$d^2(i, i') = (i - i')^{\mathrm{T}} T^{\mathrm{T}} R^{\mathrm{T}} (T T^{-1}) R T (i - i'), \tag{27}$$

where we have used the orthogonality of the transform ($T^{\mathrm{T}} = T^{-1}$), such that

$$d^2(i, i') = (T^{\mathrm{T}} R T (i - T^{\mathrm{T}}) R T i')^{\mathrm{T}} (T^{\mathrm{T}} R T (i - T)^{\mathrm{T}} R T i') \tag{28}$$

and taking into account that, in this model, the perceived image, $\hat{\imath}$, is the inverse transform of the weighted spectrum, $\hat{\imath} = T^{\mathrm{T}} R T i$, we will have:

$$d^2(i, i') = (\hat{\imath} - \hat{\imath}')^{\mathrm{T}} (\hat{\imath} - \hat{\imath}'). \tag{29}$$

In this way, the perceptual difference between two input images, i and i', is given by the euclidean difference between the filtered versions, $\hat{\imath}$ and $\hat{\imath}'$, as reported in the approaches that use qualitative reasoning [23–25].

$$d^2(i, i') = \sum_{x_i} (\hat{\imath}_{x_i} - \hat{\imath}'_{x_i})^2 = \sum_{f_i} (S_{f_i} I_{f_i} - S_{f_i} I'_{f_i})^2, \tag{30}$$

where S_{f_i} is the filter or sensitivity function, and I and I' are the expansions of the images i and i' in the selected basis.

In short, the application of the general distance definitions (Eqs. (3)–(5)) and the metric dependence with the response (Eqs. (16) and (18)) to a particular linear perception model gives the expected results: the differences in the frequency

domain must be weighted by the filter function, and the perceptual importance of the spatial surround is related to the impulse response of the system.

2.4. Perceptual metrics from complex perception models

Although in the distortion measure, proposed in this paper (Section 4), we restrict ourselves to post-transform non-linearities without inter-coefficient masking (Section 3), in this section some comments will be made about the application of the presented formulation (Eqs. (14)–(24)) to the more general contrast response models including post-transform feature interactions [6].

In the more recent contrast perception models, the cross-coefficient interactions are modeled by a normalization of the responses by a pooled signal over the channels [6].

$$R_{p_i}(I) = \frac{I_{p_i}^n}{B + \sum_{p_j} H_{p_i p_j} I_{p_j}^m} = \frac{I_{p_i}^n}{P_{p_i}}, \tag{31}$$

where I_{p_i} is the response of the Gabor filter p_i, n and m are the excitatory and inhibitory exponents (with $n > m$, and $n \sim 2$), and the inhibitory pooling is controlled by the kernel, $H_{p_i p_j}$, which gives the effect of the feature p_j in the response p_i. To obtain the metric in the feature domain, the gradient matrix of the response function has to be computed (Eq. (16)):

$$\nabla R(I)_{p_i p_j} = \frac{\mathrm{d}R_{p_i}(I)}{\mathrm{d}I_{p_j}} = n \frac{I_{p_i}^{n-1}}{P_{p_i}} \delta_{p_i p_j} - m \frac{I_{p_i}^n I_{p_j}^{m-1}}{P_{p_i}^2} H_{p_i p_j}. \tag{32}$$

Note that the gradient is input-dependent and has a positive diagonal term, δ, given by the excitatory contribution, and a negative (inhibitory) term proportional to the kernel H. As there exist interaction (masking) between coefficients, the kernel H and the corresponding metric W in the feature space are non-diagonal.

The presented formalism considers masking between features in the distortion measure as non-diagonal elements in the metric of the feature space. It can also be used to compute perceptual metrics in the spatial domain even with non-linear feature extraction transforms using the gradients of the transforms.

The main restriction of the proposed formalism is the ideal observer assumption, i.e. $\beta = 2$ in the Minkowski summation in the response domain [6]. Vectorial summation of the distortions in each dimension has to be assumed to use the tensor calculus to derive the relations between the metric in the different representations. This choice implies a restriction of the convexity of the JND regions ($\beta = 2$ implies ellipsoids), but does not restrict more important characteristics as their relative volume or orientation. The presented formulation can be applied to a wide range of contrast perception models including non-uniform sensitivity, non-linear transforms and masking between features.

In this work a physiological model of the non-linear response to the contrast of periodic patterns is proposed to be used with the presented formulation to obtain a particular distortion measure. This model and the geometric formalism provide a theoretical foundation to our previous non-linear perceptual difference measure based on the empirical bit allocation of the human visual system in a frequency domain [28].

3. Non-linear physiological model of contrast response summation

In the usual psychophysical experimentation, sinusoidal patterns of different frequency are used to measure the contrast incremental thresholds. Hence, a frequency feature space is assumed through a Fourier transform T. In this case, the strength of each feature, of each frequency f_i, is the contrast (the normalized amplitude) of the sinusoidal grating of frequency f_i.

Contrast incremental thresholds were usually measured for fixed frequencies as a function of the background contrast, giving a function, $\Delta C_{f_i}(C_{f_{ii}})$, named Contrast Discrimination Function (CDF). The CDF was described as dipper-shaped. As background contrast increases, the contrast incremental threshold decreases, giving a minimum value lower than the contrast absolute threshold when the pedestal contrast is close to the absolute threshold [15–16]. After this behavior in the low contrast range, contrast incremental threshold rises exponentially with background contrast. The slope of CDF is a log–log plot, and varies from a low value at low spatial frequencies to a value near unity at high spatial frequencies [37–38].

The physiological evidence of the existence of two main parallel pathways with different functions [17,39–41] for the information processing in the human visual system can help us to understand the observed differences. The parvocellular pathway is specially sensitive to high spatial-low temporal frequencies, giving sustained responses, whereas the magnocellular pathway drives the transient responses of human visual system to low spatial–high temporal frequencies [18,42]. A wide variety of topics relating the psychophysical experiments, which correlates with the properties of the physiological pathways, can be found in the literature [17,42–44].

A change in the linear part of the CDF is observed if the spatiotemporal characteristics of the stimuli used in the experience facilitates the response of magno or parvocellular pathways [37,45,46]. These results seem to be consistent with the hypothesis of a dual mechanism for signal processing in the human visual system, mediated by two physiological pathways, the magnocellular and the parvocellular. If such is the case, a theoretical prediction of the CDF curves may be derived from the contrast response functions (CRF) of the cells of these pathways.

Here, we propose a frequency-dependent model of response summation of the cells of the parvocellular

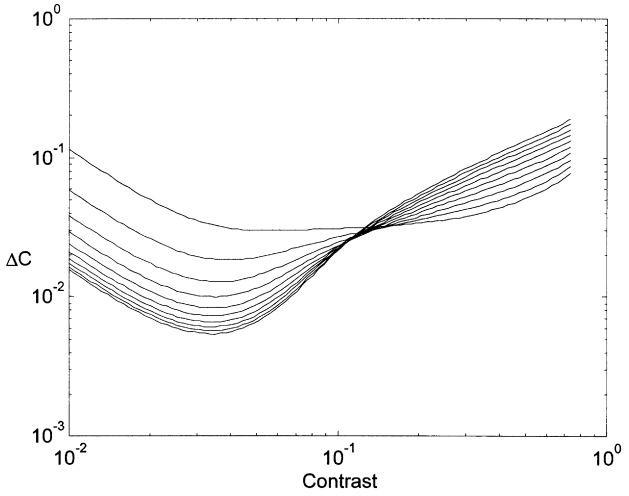

Fig. 4. CDFs inferred by the variation of the summation index.

pathway to reproduce the psychophysical contrast discrimination data of each global frequency detection mechanism.

The contrast response function of an individual cell of the parallel pathways can be fitted as a Naka–Rushton function [47–48]:

$$R_{f_i,\alpha}(C_{f_i}) = R^0_{f_i}\frac{C^n_{f_i}}{C^n_{f_i}+\alpha^n_{f_i}}, \tag{33}$$

where $R^0_{f_i}$ is a constant related to the saturation level of the cell, α_{f_i} is the semisaturation constant and the exponent n is the parameter that controls the slope of the CDF.

Nevertheless, the extensive works of Albretch and Hamilton [47] and Sclar et al. [48] studying the contrast response of cells at different stages of the visual system, give us a set of fitting parameters of the response function. These parameters show significant statistical differences between the different cells in a pathway, specially α_{f_i}, the parameter with more distortion. This implies that the global contrast response of the system cannot be derived from the response

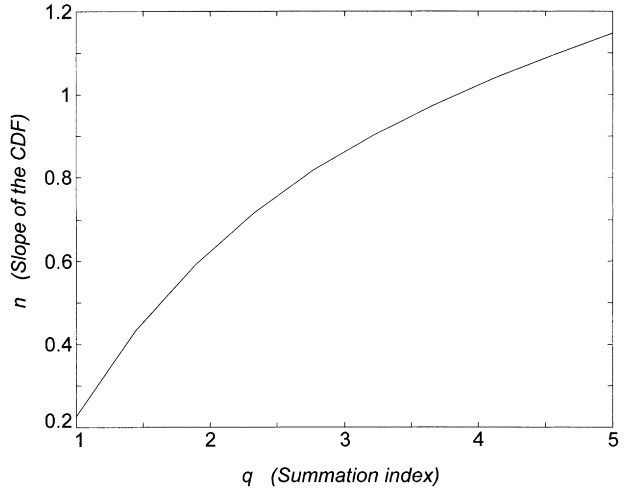

Fig. 5. Relationship between the slope of the CDF and the summation index, q.

expression of a single cell. In this way, a contrast response summation model must be proposed to collect all the contributions from each cell to give a global response that reproduces the CDF data.

The experiences of small lesions in the LGN in monkeys, affecting the magno or parvo pathways suggests that the contrast pattern processing is mediated by the parvocellular pathways, whereas the magnocellular pathway has only a minor contribution [40–41]. The role of the magnocellular pathway is to mediate the temporal aspects of perception, such as motion and high frequency flicker detection [18].

If this be the case, it is possible to suggest that the discrimination task can be carried out fundamentally with the parvocellular contributions. The change in the slope of the CDF would then be related to the spatiotemporal sensitivity of this pathway. To test this hypothesis, we can assume that the different, α_{f_i}-dependent cell responses are combined by probability summation, to yield the global response, $R_{f_i}(C_{f_i})$. This global response should control the performance of the detection mechanism of frequency f_i [30]:

$$R_{f_i}(C_{f_i}) = \left(\sum_{\alpha}(R_{f_i,\alpha}(C_{f_i}))^q\right)^{1/q} \tag{34}$$

where q is the probability summation index.

This expression is similar to Eq. (2), which combines the responses of the different feature detection mechanisms, but note that Eq. (34) describes the combination of different response cells pertaining to a single frequency detection mechanism. Taking the α_{f_i} distribution values for the parvo pathway, we can calculate the combined response of the cells at V1 level for different values of q (from $q=1$ to $q=5$). The contrast incremental thresholds can be derived from Eqs. (13) and (34), yielding

$$\Delta C_{f_i}(C_{f_i}) = \frac{n\Delta r_{f_i}}{R^0_{f_i}}C^{-n-1}_{f_i}\left(\sum_{\alpha}R_{f_i,\alpha}(C_{f_i})^q\right)^{(1/q)-1} \times\left(\sum_{\alpha}R_{f_i,\alpha}(C_{f_i})^{q+1}\alpha^n\right), \tag{35}$$

where Δr_{f_i} is the response incremental threshold in the neural representation, which has been assumed to be a constant for every frequency and amplitude.

If we derive the CDF from the response curves, we obtain curves that reproduce the behavior of the CDF even in the low-contrast range. A linear behavior in the medium–high contrast range (above 10%) is observed, with the slope increasing with the value of q. Fig. 4 shows the set of CDFs generated for different q values. Fig. 5 plots the variation of the slope of the linear part of the CDF with the probability summation constant. We observe that for values of high probability summation ($q \approx 1$), the exponent of the CDFs have a value of 0.4. The slope reaches unity when q increases to $q \approx 4$.

These psychophysical results suggest that only one mechanism (with spatiotemporal sensitivities similar to

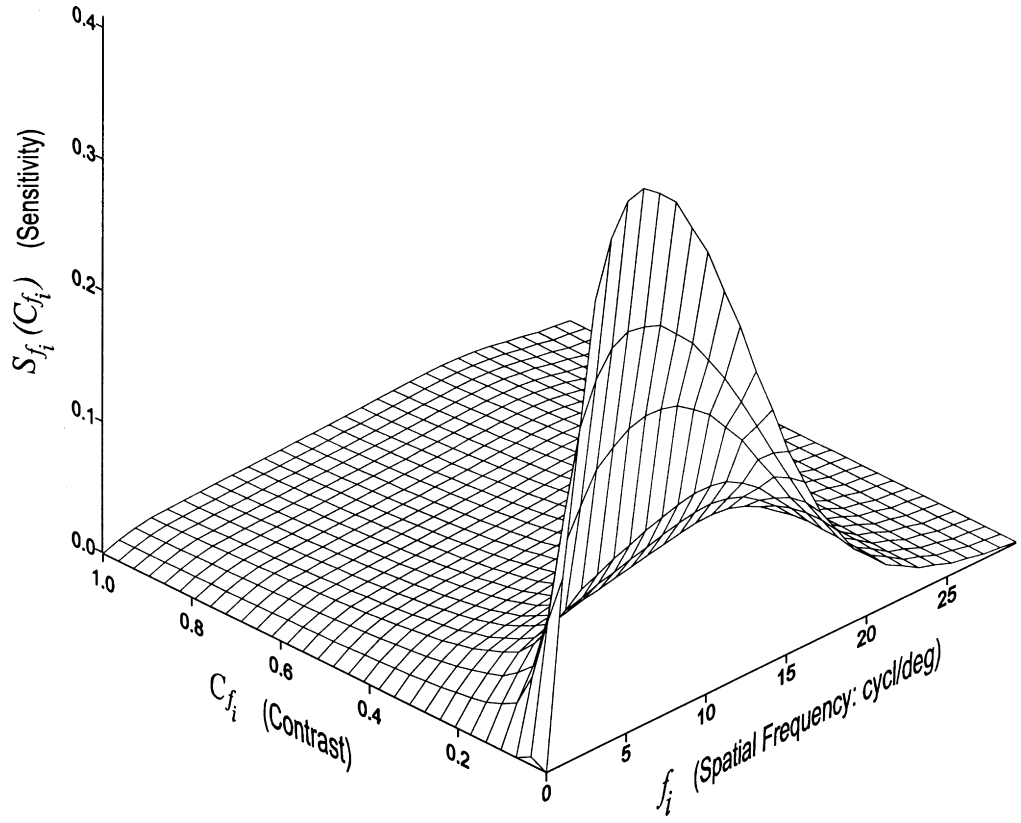

Fig. 6. Sensitivity function for every frequency and contrast value. It can be seen how the $S_{f_i}(I_{f_i})$ reproduces the CSF when the contrast tends to zero. The amplitude non-uniformities of sensitivity are no longer simple exponential decreases as it would assume contrast power incremental thresholds [15].

that of the parvocellular pathway) drives the response of the system for contrast discrimination. When the spatio-temporal characteristics of the stimulus are not adequate for this mechanism, i.e. short duration stimuli or low spatial frequencies, the response is obtained from the linear summation of the pooled individual responses. Stimuli whose characteristics allow a good response from the parvocellular pathway, give a response with low probability summation. In this case, the response of the system is mediated by the most sensitive cells to the stimuli, and a slope close to unity is obtained. Although vector models of interaction ($q = 2$) are the most common, values of q around 4 have been proposed by other authors [49].

The variation of the constant q can be interpreted as an actual gain control mechanism, by varying the combined response curve as a function of the input stimulus in order to provide a reliable response. This model can explain in a simple way the variation of the slope of the CDF with spatial frequency, in contrast with previous models of contrast discrimination. From Fig. 5, we will obtain an empirical expression of the relationship, $q = q(f_i)$:

$$q_{f_i} = q_0\left(\frac{f_i^a}{f_i^q + b}\right) \tag{36}$$

q_0, a and b being the best fitting constants.

4. Non-linear sensitivity and subjective distortion in a frequency domain

In this section, a novel distortion measure is derived from the presented formalism particularized through the proposed contrast response summation model.

This distortion measure weighs the differences between the amplitudes of the feature representation coefficients by a non-uniform sensitivity function related to the number of discriminable patterns per unit of area in the frequency and contrast domain: the *Information Allocation Function* [28,50].

As in the proposed model, the response to each frequency pattern is not affected by any other basis pattern in the input image, the linear approximation of R, the gradient ∇R is a point-dependent diagonal matrix. The elements of the diagonal of $\nabla R(I)$ are the slopes of the unidimensional response curves $R_{f_i}(I_{f_i})$, or $R_{f_i}(C_{f_i})$, at the normalized amplitude I_{f_i} (contrast C_{f_i}) for each coefficient f_i.

Sensitivity functions in the spatial and in the frequency domain can be derived from the response model (Eqs. (34) and (36)) and from the relations (20) and (21). In this case, a sensitivity value can be obtained for each 2D pattern.

In the same way, the sensitivity in each dimension through the expressions (23) and (24) can also be obtained.

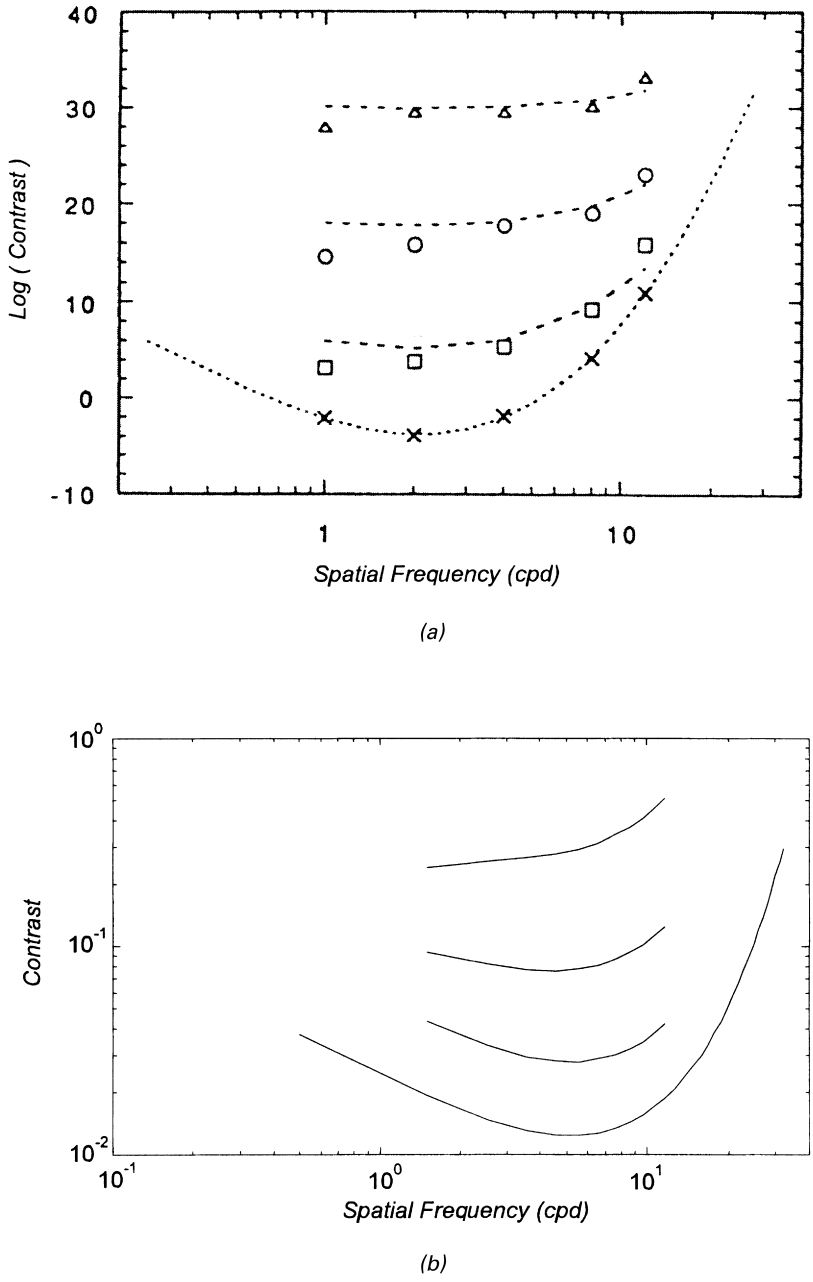

Fig. 7. (a) Curves of equal perceived contrast of Georgeson et al. [19], (b) equivalent results generated by the proposed model.

As the gradient is diagonal, only one term of the sum in the expression (23) is non-null, so the sensitivity for each transform coefficient as a function of the amplitude is just the slope of the corresponding unidimensional response curve at that amplitude point, $S_{f_i}(I_{f_i}) = \partial R_{f_i}(I_{f_i})/\partial I_{f_i}$. The sensitivity at each position is more difficult to analyze because the transform T mixes the contributions of all the transform coefficients for every spatial position.

Fig. 6 shows the sensitivity function for each dimension of the frequency feature domain at different levels of contrast for each coefficient. The contrast non-linearities of the response model imply the non-uniformity of the sensitivity for every coefficient. The particular result shown here, the decrease of the incremental thresholds in the low contrast region for every frequency has been approximated to a constant value equal to the absolute threshold as in Refs. [5,50], so the actual sensitivity surface should be slightly higher in the low, over threshold, contrast region.

As stated above, the sensitivity is proportional to the number of discriminable patterns of a given frequency per unit of amplitude (contrast). The non-uniformity in the resolution shown in Fig. 6 was interpreted as a perceptual information allocation function.

A system that distinguishes n different values of a signal

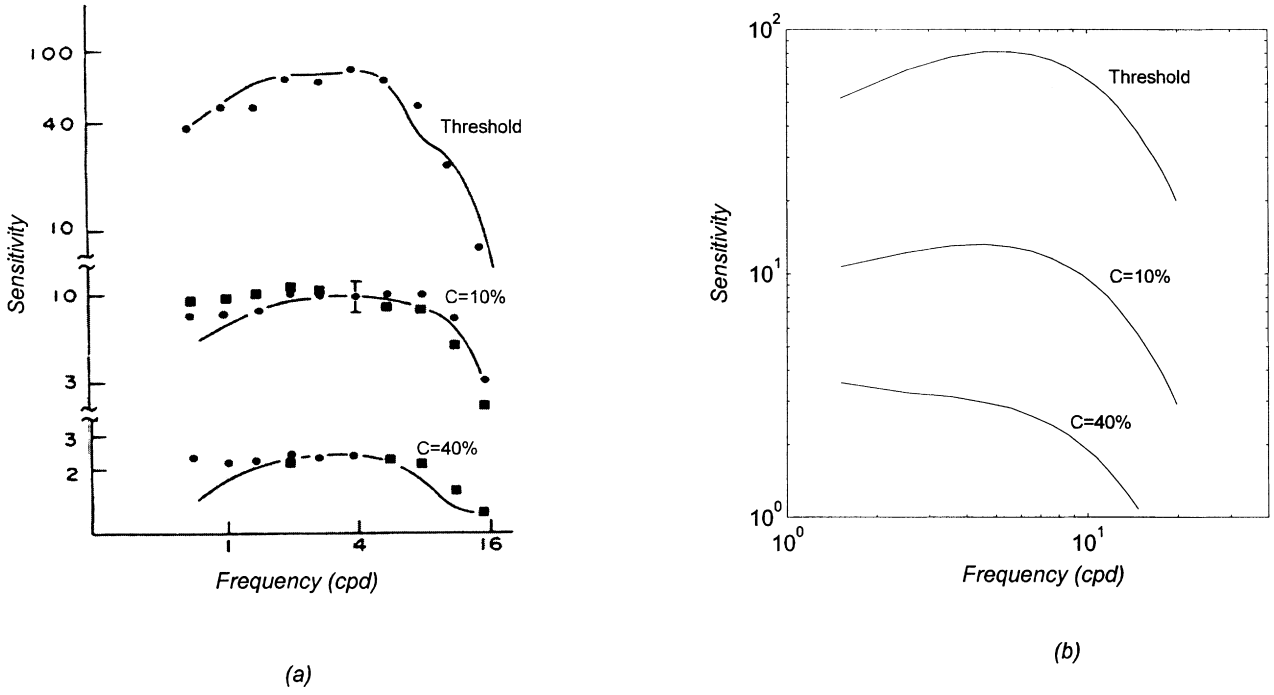

Fig. 8. (a) Contrast matching results of Swanson et al. [20] at threshold contrast, 10 and 40% contrast, (b) equivalent results generated by the proposed model.

is said to use $\log_2(n)$ bits to encode the signal [51]. Hence, if the sensitivity $S_{f_i}(I_{f_i})$ is proportional to the number of discriminable patterns per unit of contrast, the bits used by the system to encode each region of the frequency and contrast domain will be related to the logarithm of the sensitivity function. This is the reason why the sensitivity function $S_{f_i}(I_{f_i})$ has been called *Information Allocation* (IAF) of the human visual system [28,50]. It has been used to design a perceptually matched quantizer for image and video compression [50,52]. The IAF has also been used to design a perceptual distortion metric to weight the differences between the amplitudes of a quantized *Cosine Transform* of the input images [28]. In this case, qualitative reasonings were used to construct the difference expression.

Now, a strictly derived expression is obtained from Eqs. (4) and (16), and the model of contrast response. As $\nabla R(I)$ is diagonal, $W(I)$ will also be diagonal, and the metric element corresponding to each frequency component will be $W(I)^{f_i}_{f_i} = [S_{f_i}(I_{f_i})]^2$. So, given two images i and $i + \Delta i$, with frequency transforms I and $I + \Delta I$, the perceptual difference between them will be given by:

$$D^2(I, I + \Delta I) = \Delta I^{\mathrm{T}} W \Delta I = \sum_{f_i} [S_{f_i}(I_{f_i})]^2 \Delta I^2_{f_i}. \tag{37}$$

5. Suprathreshold contrast and distortion measurement results

In this section, the proposed model for contrast perception and image distortion evaluation is tested. On the one hand, it is shown how the response model reproduces different experimental suprathreshold contrast results [19–22]. The predicted behavior is closer to the experimental data than in the models with frequency-independent summation index at cell level. On the other hand, it is shown how the distortion prediction evaluated through the proposed metric is linearly related to the observer's opinion under different noise conditions.

In this work we focus on the fit of experimental results by the proposed model and on the analysis of the quality of such predictions, but not on the experimental details. The interested reader is referred to the specific references where the experimental methods are described.

5.1. Suprathreshold contrast results from the contrast response summation model

Assuming that the response of the mechanism to a certain contrast, $r_{f_i}(C_{f_i})$, can be related directly to the perceived contrast [21,53], the classic experiences of suprathreshold perceived contrast should be reproduced by the proposed model.

Fixing a given response value, equal response contrast curves can be obtained looking for the pair (f_i, C_f) that satisfied Eq. (34) with a frequency-dependent q (Eq. (36)).

Fig. 7(b) represents the equal perceived contrast curves as a function of the spatial frequency derived in this way. The obtained result is similar to the experimental one obtained by Georgeson and Shackleton [19], Fig. 7(a), also reported by Bowker [54].

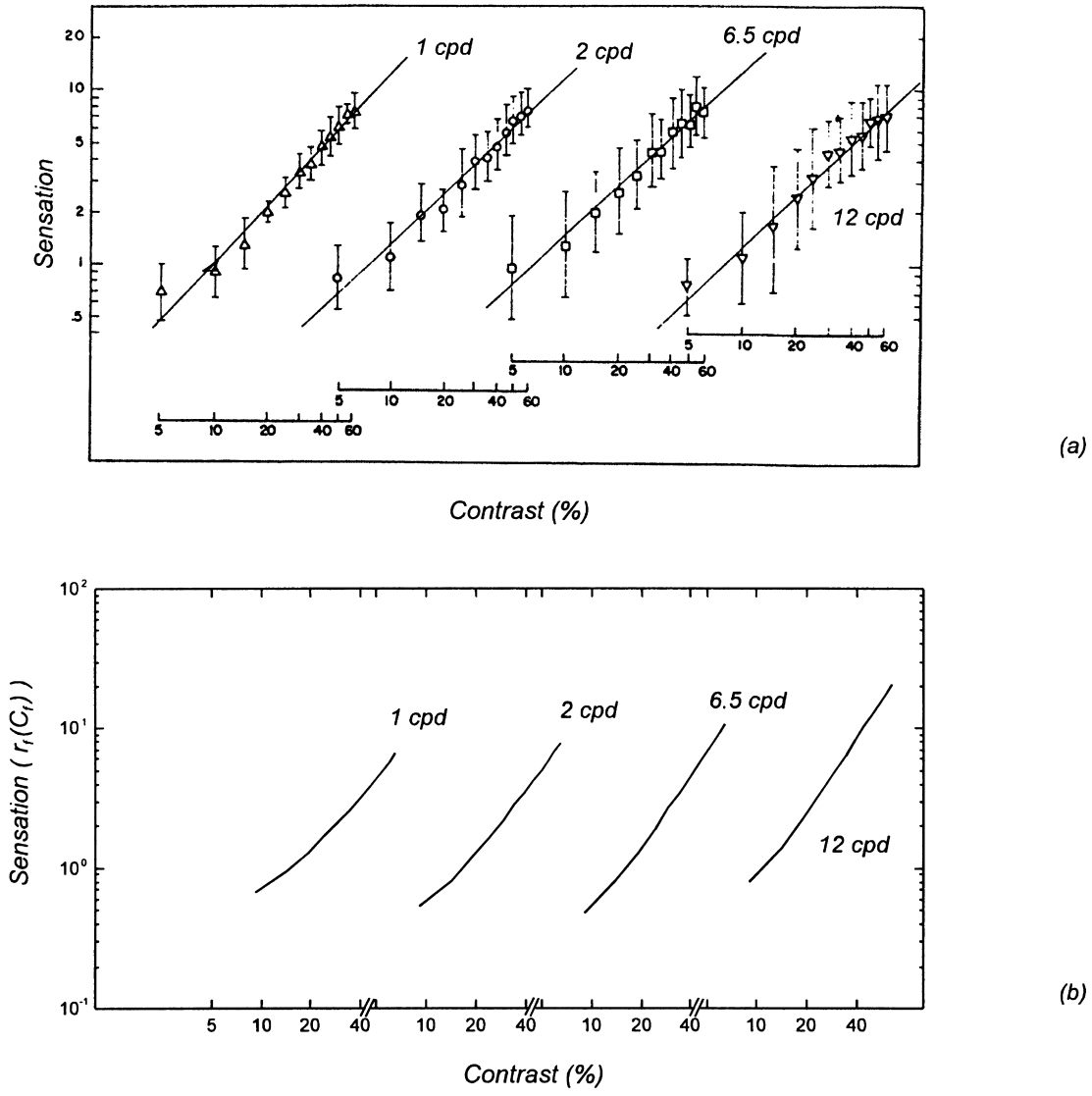

Fig. 9. (a) Perceived contrast results of Cannon [21] as a function of the input contrast for several frequencies, (b) equivalent results generated by the proposed model.

The proposed model also reproduces the experimental results of suprathreshold sensitivity measured through contrast matching. For instance, our frequency-dependent response summation model achieves a better reproduction of the experimental results of Swanson et al. [20] than the model proposed by these authors, specially at the low frequencies region. It is interesting to point out that in this work they use a fixed, frequency-independent, summation index.

Fig. 8(a) shows the results of these authors together with their predictions (continuous line in the figure). It can be observed that at high contrasts, the low spatial frequencies are not adjusted by their model. In our model, we consider the possibility to vary the summation index with the spatial frequency. When introducing this possibility, we can also reproduce the results of these authors in the low spatial frequency range (Fig. 8(b)).

The method of numerical estimation of the magnitude developed by Stevens [55] is, potentially, the best way to relate a given contrast and the sensation that it produces (perceived contrast). In this method, the observer should assign a numeric value to a series of contrasts, building a scale of perceived contrasts. It is widely accepted that a linear relationship between the perceived contrast and the contrast of a sinusoidal grating is found [21–22,53]. This result is reproduced by the presented model, which plots the response curves as a function of the input contrast. Fig. 9 represents the results of estimation of the magnitude of Cannon [21] compared with those generated from our model. Note how the actual behavior of the experimental data is non-linear in this representation as predicted by the model.

It is also possible to compare our model predictions with the curves of equal contrasts obtained by Biondini and

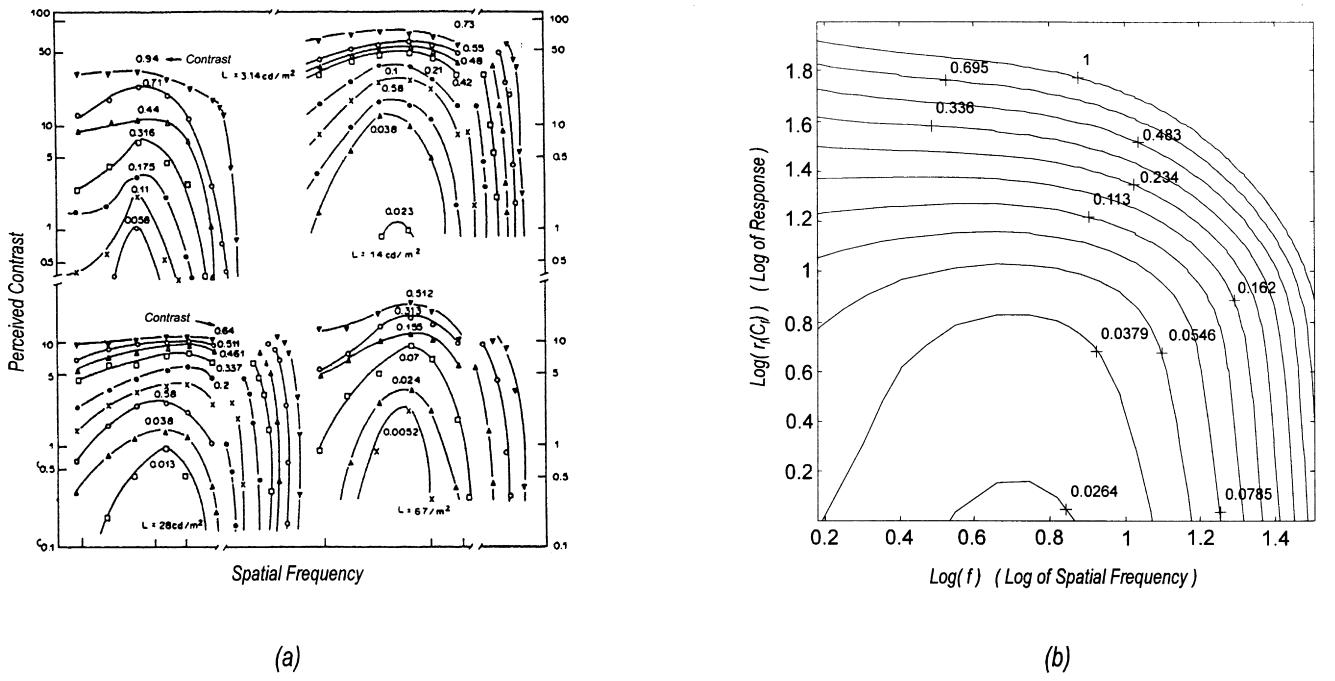

Fig. 10. (a) Curves of equal input contrast in a perceived contrast-frequency plane for several luminances [22], (b) equivalent results generated by the proposed model. Our model does not include background luminance effects.

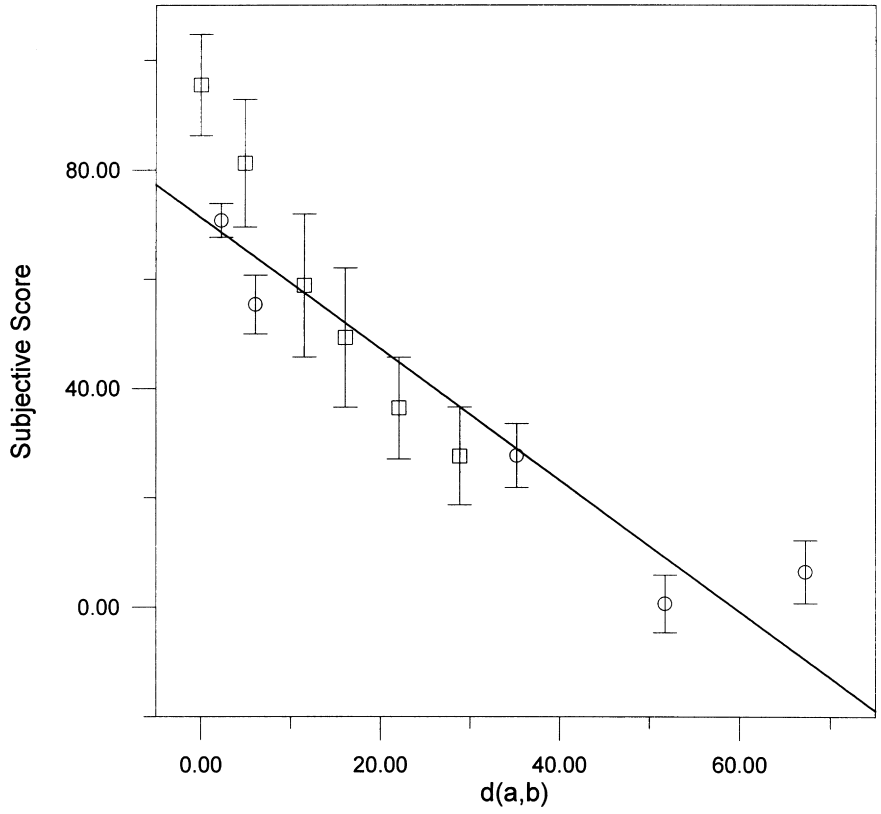

Fig. 11. Perceptual quality judgements in a variety of noise conditions [28] versus the difference value computed by Eq. (35).

Mattiello [22] in their exhaustive study on the relationship between the perceived contrast and the contrast of the stimulus. They represent curves of equal contrasts of the stimulus in the plane of frequencies and perceived contrast for different mean luminances (Fig. 10(a)). These curves represent the differences of perceived contrast as a function of the spatial frequency for a given input contrast. Fig. 10(b) shows the great qualitative coincidence among the experimental curves and the predictions of the model.

As can be seen, our model reproduces the main results of the contrast perception given in the bibliography. The consideration of a variation of the summation index with the spatial frequency supposes an improvement of previously proposed models, so the use of this model to evaluate the perceptual effect of contrast differences should give a reliable subjective difference measure.

5.2. Subjective quality assessment from the contrast response summation model

In this section, the experimental data about subjective evaluation of distortion is linearly fitted to the dissimilarity predicted by the algorithm. The results of two different experiments are successfully reproduced by the same expression, showing the reliability of the proposed metric. The experimental data come from standard psychophysical procedure used elsewhere [27], in which a set of observers evaluated the distortion of a natural image corrupted by: (a) equal energy noise with different spectrum, and (b) JPEG-like lossy compression at different compression ratios. These experiments are described in detail in Ref. [5].

Fig. 11 displays the quality assessment of the observers (normalized between 0 and 100) plotted versus the numerical results given by Eq. (37) in experiment: (a) circles, and (b) squares. It used the approximated sensitivity surface of Fig. 4 from Refs. [5,50].

A high correlation coefficient is obtained, $\rho = 0.93$, and the χ^2 test value (over 0.18) confirms the goodness of the simultaneous linear fit of both experiments [33]. The quality of the fit improves the results of several previously reported metrics (Non-linear MSE [27,56], simple masking based metrics through variance and local luminance [27,56], Subjective Quality Factor and Square Root integral method [25]) and is comparable to CSF-based metrics [23,25,27], see the corresponding results in Ref. [5]). This preliminary result, using the approximated frequency and amplitude sensitivity surface, does not achieve as good results as our IAF-based metric in Ref. [5]. In this case, a correlation coefficient of 0.98 and a χ^2 test value of 0.9 was obtained for the same experimental data by the heuristically derived expression:

$$D^2(I, I') = \sum_{f_i} S_{f_i}(I_{f_i})(QI - QI')^2, \tag{38}$$

where QI and QI' are the quantized versions of the transform representations, I and I', assuming a quantizer with bit allocation given by the IAF, or $S_{f_i}(I_{f_i})$.

The rationale of this expression is to consider perception as a representation process in a quantized domain (with discrete perceptions). The density of such discrete perceptions in the feature space should be proportional to the number of discriminable patterns per unit of area, i.e. proportional to the sensitivity. In this way, the perceptual difference should be computed from the IAF *quantized* images and the remaining differences should also be weighted by the IAF.

Expressions (37) and (38) are qualitatively similar. In both the cases, the difference measure results from applying the IAF twice. On the one hand in the heuristically derived case [28], it is first applied to quantize the images and then the resulting differences are weighted by the IAF. And on the other hand, in the strictly derived expression presented here (Eq. (37)), its square is applied to weight the transform differences.

6. Final remarks

In this paper, a formalism to obtain the subjective metric in the spatial or in the feature domain from the visual system response was presented. These relations are general enough to be used in complex human visual system model. They deal with non-linear feature extraction schemes and masking between the features. The result is an image-dependent metric determined by the non-linearities of the feature extraction transform and the non-linearities of the mechanisms tuned to each feature.

Assuming a certain feature extraction transform T, explicit expressions are given to obtain the metric from experimental data about amplitude incremental thresholds of the basis patterns of T, or from a model of the system response to the basis of T.

This general formalism gives rise to some previously reported metrics when a simple visual processing model is assumed [23–27]. Here a novel perceptual difference measure is proposed particularizing the general expressions with a perception model that assumes a frequency-based transform and a non-linear post-transform transduction. The experimental non-linearities of the response functions are derived from a frequency-dependent model of response summation of the cells of the parvocellular pathway.

The model accurately reproduces the results of different suprathreshold contrast experiments. The distortions predicted by a preliminary version of the proposed metric are linearly related with experimental distortion appearance under a variety of noise conditions. An appropriate consideration of the increase in the sensitivity in near threshold contrast should be taken into account to get a more accurate distortion evaluation results.

Better subjective metrics can be obtained particularizing the presented formalism with models that account for

interchannel masking [6], as indicated in Section 2.4. Future research should be oriented to compare these refined versions of the proposed algorithm with the recently developed metrics [57–59].

Acknowledgements

This work has been partially supported by the projects CICYT 1FD 97-0279(TIC) and IVEI 96/003-35.

References

[1] N. Jayant, J. Johnston, R. Safranek, Signal compression based on models of human perception, Proc. IEEE 81 (1993) 1385–1422.
[2] M.D. Fairchild, Color Appearance Models, Addison Wesley, Reading, MA, 1977 Chapter 10.
[3] H.R. Wilson, D. Levi, L. Maffei, J. Rovamo, R. DeValois, The perception of form: retina to striate cortex, in: L. Spillmann, J.S. Werner (Eds.), Visual perception: the Neurophysiological Foundations, Academic Press, San Diego, 1990 Chapter 10.
[4] J. Lubin, The use of psychophysical data and models in the analysis od display system performance, in: A.B. Watson (Ed.), Digital Images and Human Vision, MIT Press, Cambridge, MA, 1993 pp. 163–178.
[5] S. Daly, Visible differences predictor: an algorithm for the assessment of image fidelity, in: A.B. Watson (Ed.), Digital Images and Human Vision, MIT Press, Cambridge, MA, 1993 pp. 179–206.
[6] A.B. Watson, A model of visual contrast gain control and pattern masking, J. Opt. Soc. Am. A 14 (1997) 2379–2391.
[7] A.B. Watson, The cortex transform: rapid computation of simulated neural images, Comp. Vis. Graph. Im. Proc. 39 (1987) 311–327.
[8] J.G. Daugman, Spatial visual channels in the Fourier plane, Vision Res. 24 (1984) 891–910.
[9] H.R. Wilson, J. Gelb, Modified line-element theory for spatial-frequency and width discrimination, J. Opt. Soc. Am A. 1 (1984) 124–131.
[10] B.A. Wandell, Foundations of vision, *Multiresolution Image Representations*, Sinauer Associates, Massachusetts, 1995 Chapter 8.
[11] J. Malo, A. Felipe, A.M. Pons, J.M. Artigas, Characterization of the human visual system threshold performance by a weighting function in the Gabor domain, J. Mod. Opt. 44 (1997) 127–148.
[12] A.C. Bovik, M. Clark, W.S. Geisler, Multichannel texture analysis using localized spatial filters, IEEE Trans. Patt. Anal. Machine Intell. 12 (1990) 55–73.
[13] M.S. Landy, J.R. Bergen, Texture segregation and orientation gradient, Vision Res. 31 (1991) 679–691.
[14] J. Nachmias, R.V. Sansbury, Grating contrast: discrimination may be better than detection, Vision Res. 14 (1974) 1039–1042.
[15] G.E. Legge, A power law for contrast discrimination, Vision Res. 18 (1981) 68–91.
[16] G.E. Legge, J.M. Foley, Contrast masking in human vision, J. Opt. Soc. Am. 70 (1980) 1458–1471.
[17] P. Lennie, Parallel Visual pathways: a review, Vision Res. 20 (1980) 561–594.
[18] A.M. Derrington, P. Lennie, Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque, J. Physiol. 357 (1984) 291–340.
[19] M.A. Georgeson, T.M. Shackleton, Perceived contrast of gratings and plaids: nonlinear summation across oriented filters, Vision Res. 34 (1994) 1061–1075.
[20] W.H. Swanson, H.R. Wilson, S.C. Giese, Contrast matching data inferred from contrast incremental thresholds, Vision Res. 24 (1985) 63–75.
[21] M.W. Cannon, Contrast sensation: a linear function of stimulus contrast, Vision Res. 19 (1979) 1045–1052.
[22] A.R. Biondini, M.L.F. Mattiello, Suprathreshold contrast perception at different luminance levels, Vision Res. 25 (1985) 1–9.
[23] N.B. Nill, B.R. Bouzas, Objective image quality measure derived from digital image power spectra, Opt. Engng 32 (1992) 813–825.
[24] E.M. Granger, K.N. Cupery, An optical merit function (SQF) which correlates with subjective image judgements, Photog. Sci. Engng 16 (1972) 221–230.
[25] P.J.G. Barten, Evaluation of subjective image quality with the SQRI method, J. Opt. Soc. Am. A 7 (1990) 2024–2031.
[26] J.A. Saghri, P.S. Cheatham, A. Habibi, Image quality measure based on a human visual system model, Opt. Engng 28 (1989) 813–818.
[27] D.R. Fuhrmann, J.A. Baro, J.R. Cox, Experimental evaluation of psychophysical distortion metrics for JPEG-encoded images, Proc. SPIE (1993) 179–190.
[28] J. Malo, A.M. Pons, J.M. Artigas, Subjective image fidelity metric based on bit allocation of the human visual system in the DCT domain, Im. Vis. Comp. 15 (1997) 535–548.
[29] A.B. Watson, Detection and recognition of spatial forms, in: Braddick, Sleigh (Eds.), Physical and Biological Processing of Images, Springer, Berlin, 1983 pp. 100–114.
[30] R.F. Quick, A vector magnitude model of contrast detection, Kybernetic 16 (1974) 65–67.
[31] A.H. Barr, Superquadrics and angle-preserving transformations, IEEE Comp. Graph. Appl. 1 (1981) 11–23.
[32] D.H. Kelly, Receptive field like functions inferred from large area psychophysical measurements, Vision Res. 25 (1985) 1895–1900.
[33] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C: the Art of Scientific Computing, Cambridge University Press, Cambridge, 1992.
[34] P. Capilla, J. Malo, M.J. Luque, J.M. Artigas, Colour representation spaces at different physiological levels: a comparative analysis, J. Opt. 29 (1998) 324–338.
[35] B.A. Dubrovin, A.T. Fomenko, S.P. Novikov, Modern geometry: methods and applications, *Tensors: the Algebraic Theory*, Springer, New York, 1984 Chapter 3.
[36] A.N. Akansu, R.A. Haddad, Multiresolution signal decomposition, *Orthogonal Transforms*, Academic Press, Boston, 1992 Chapter 2.
[37] E.L. Smith III, R.S. Harwerth, D.M. Levi, R.L. Boltz, Contrast increment thresholds of rhesus monkeys, Vision Res. 22 (1982) 1153–1161.
[38] A.M. Pons, P. Capilla, J. Malo, A. Felipe, J.M. Artigas, Contrast discrimination function and the transient sustained dichotomy, Perception 24 (1995) 89.
[39] D.H. Hubel, M.S. Livingstone, Color and contrast sensitivity in the lateral geniculate body and perimary visual cortex of the macaque monkey, J. Neurosci. 10 (1990) 2223–2237.
[40] W.H. Merigan, J.H.R. Maunsell, Macaque vision after magnocellular lateral geniculate lessions, Visual Neurosci. 5 (1990) 347–352.
[41] W.H. Merigan, L.M. Katz, J.H.R. Maunsell, The effects of parvocellular lateral geniculate lesions on the acuity and contrast sensitivity of macaque monkeys, J. Neurosci. 11 (1991) 994–1001.
[42] J.J. Kulikowski, D.J. Tolhurst, Psychophysical evidence for sustained and transient detectors in human vision, J. Physiol. 232 (1973) 149–163.
[43] C.R. Ingling, S.S. Grigsby, Perceptual correlates of magnocellular and parvocellular channels: seeing form and depth in afterimages, Vision Res. 30 (1990) 823–828.
[44] B.B. Lee, J. Pokorny, V.C. Smith, P.R. Martin, A. Valberg, Luminance and chromatic modulation sensitivity of macaque ganglion cells and human observers, J. Opt. Soc. Am. A 7 (1990) 2223–2236.
[45] T.B. Lawton, C.W. Tyler, On the role of X and simple cells in human contrast processing, Vision Res. 34 (1994) 659–667.
[46] J.I. Nelson, W.H. Seiple, Human VEP contrast modulation sensitivity: separation of magno- and parvocellular components, Electroencelography Clin. Neurophysiol. 84 (1992) 1–12.

[47] D.G. Albretch, D.B. Hamilton, Striate cortex of monkey and cat: contrast response function, J. Neurophysiol. 48 (1982) 217–237.
[48] G. Sclar, J.H.R. Maunsell, P. Lennie, Coding of image contrast in central visual pathways of the macaque monkey, Vision Res. 30 (1990) 1–10.
[49] K. Kranda, P.E. King-Smith, Detection of colored stimuli by independent linear systems, Vision Res. 19 (1979) 733–745.
[50] J. Malo, A.M. Pons, J.M. Artigas, Bit allocation algorithm for codebook design vector quantization fully based on human visual system non-linearities for suprathreshold contrasts, Electr. Lett. 31 (1995) 1229–1231.
[51] N. Abramson, Information theory and coding, McGraw-Hill, New York, 1964.
[52] J. Malo, F. Ferri, J. Albert, J.M. Artigas, Splitting criterion for hierarchical motion estimation based on perceptual coding, Elect. Lett. 34 (1998) 541–543.
[53] M.W. Cannon, S.C. Fullemkamp, A transducer model for contrast perception, Vision Res. 31 (1991) 983–989.
[54] D.O. Bowker, Suprathreshold spatiotemporal response characteristics of the human visual system, J. Opt. Soc. Am 73 (1985) 436–440.
[55] S.S. Stevens, On the psychophysical law, Psychol. Rev. 64 (1957) 153–181.
[56] H. Marmolin, Subjective MSE measures, IEEE Trans. Sys. Man Cyber. 16 (1986) 486–489.
[57] A.B. Watson, R. Borthwick, M. Taylor, Image quality and entropy masking, Proc. SPIE 00 (1997) 3016.
[58] V. Kayargadde, J.B. Martens, Perceptual characterization of images degraded by blur and noise: model and experiments, J. Opt. Soc. Am. A 13 (1996) 1166–1188.
[59] J.M. Baena, A. Toet, X.R. Fdez-Valdivia, A. Garrido, R. Rodriguez, A computational visual distinctness metric, Opt. Engng 37 (1998) 1995–2205.